

基于 3D 人脸重建的光照、姿态不变人脸识别*

柴秀娟¹⁺, 山世光², 卿来云², 陈熙霖², 高文^{1,2}

¹(哈尔滨工业大学 计算机学院, 黑龙江 哈尔滨 150001)

²(中国科学院 计算技术研究所 ICT-ISVISION 面像识别联合实验室, 北京 100080)

Pose and Illumination Invariant Face Recognition Based on 3D Face Reconstruction

CHAI Xiu-Juan¹⁺, SHAN Shi-Guang², QING Lai-Yun², CHEN Xi-Lin², GAO Wen^{1,2}

¹(Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

²(ICT-ISVISION Joint R&D Laboratory for Face Recognition, Institute of Computer Technology, The Chinese Academy of Sciences, Beijing 100080, China)

+ Corresponding author: Phn: +86-10-58858300 ext 314, Fax: +86-10-58858301, E-mail: xjchai@jdl.ac.cn, <http://www.jdl.ac.cn/>

Chai XJ, Shan SG, Qing LY, Chen XL, Gao W. Pose and illumination invariant face recognition based on 3D face reconstruction. *Journal of Software*, 2006,17(3):525- 534. <http://www.jos.org.cn/1000-9825/17/525.htm>

Abstract: Pose and illumination changings from picture to picture are two main barriers toward full automatic face recognition. In this paper, a novel method to handle both pose and lighting conditions simultaneously is proposed, which calibrates the pose and lighting to a predefined reference condition through an illumination invariant 3D face reconstruction. First, some located facial landmarks and a priori statistical deformable 3D model are used to recover an elaborate 3D shape. Based on the recovered 3D shape, the “texture image” calibrated to a standard illumination is generated by spherical harmonics ratio image and finally the illumination independent 3D face is reconstructed completely. The proposed method combines the strength of statistical deformable model to describe the shape information and the compact representations of the illumination in spherical frequency space, and handles both the pose and illumination variation simultaneously. This algorithm can be used to synthesize virtual views of a given face image and enhance the performance of face recognition. Experimental results on CMU PIE database show that this method can significantly improve the accuracy of the existing face recognition method when pose and illumination are inconsistent between gallery and probe sets.

Key words: face recognition; 3D face reconstruction; statistic deformable model; spherical harmonic; ratio image

摘要: 待匹配人脸图像与库存原型图像之间姿态和光照的差异是自动人脸识别的两个主要瓶颈问题,已有的解决方法往往只能单独处理二者之一,而不能同时处理光照和姿态问题.提出了一种对人脸图像中的姿态和光照变化同时进行校正处理的方法,即通过光照不变的 3D 人脸重建过程,将姿态和光照都校正到预先定义的标准条件下.首先,利用先验的统计变形模型,结合人脸图像上的一些关键点来恢复较为精细的人脸 3D 形状.基于此重建的 3D 形

* Supported by the National Natural Science Foundation of China under Grant No.60332010 (国家自然科学基金); the “100 Talents Program” of the CAS (中国科学院百人计划); the Shanghai Municipal Sciences and Technology Committee of China under Grant No.03DZ15013 (上海市科委项目); the ISVISION Technologies Co., Ltd (银晨智能识别科技有限公司资金资助)

状,进而通过球面谐波商图像的方法估计输入图像的光照属性并提取输入图像的光照无关的纹理信息,从而将光照无关的 3D 人脸完全重构出来,生成输入人脸图像在标准姿态和光照条件下的虚拟视图,用于最终的分类识别,实现了对光照和姿态问题的同时处理.在 CMU PIE 数据库上的实验结果表明,此方法可以在很大程度上提高现有人脸识别方法对于原型集合(gallery)和测试集中图像在姿态和光照不一致情况下识别结果的正确性.

关键词: 人脸识别;3D 人脸重建;统计变形模型;球面谐波;商图像

中图分类号: TP393 文献标识码: A

人脸识别技术在安全、金融、法律、人机交互等领域具有广阔的应用前景,因此得到了研究人员的广泛关注.经过近 40 年的发展,对于均匀光照下的中性表情的正面人脸图像其识别率已经很高^[1].然而在一些更复杂的情况下,现有多数系统的识别性能受姿态和光照变化的影响特别大.这是因为当人脸的姿态或光照发生变化时,人脸图像的外观也会随之发生很大变化,所以通常使用的基于 2D 外观的方法在这种情况下就会失效.尽管也有一些基于 2D 的方法(如基于多视图的方法^[2])可以在一定程度上解决姿态或者光照的变化问题,但我们认为基于 3D 信息来改善 2D 图像外观的方法是解决姿态、光照变化问题的最本质的方法.

在研究的早期阶段,无论是针对姿态问题,还是光照问题,基于人脸图像外观低维子空间描述的方法都是主要的思路.Eigenfaces^[2]和 Fisherfaces^[3]根据统计学习得到经验的人脸低维的姿态或光照空间.这类方法易于实现,精度较高.但是当测试图像和训练图像集合的成像条件不太相似时,其性能下降得非常严重.由 Gross 等人提出的 Fisher Light-Fields 算法^[4]通过估计 gallery 或者测试图像对应的头部的特征光场,并将特征光场系数作为特征集合来进行最终的识别.将此工作进一步扩展,Zhou 提出了一种 Illuminating Light Field 方法^[5],其中,Lambertian 反射模型被用于控制光照的变化,该方法对新的光照条件比 Fisher Light Field 具有更强的泛化能力.然而,该算法需要多个姿态以及多种光照条件下的多幅训练图像进行建模,这对多数实际应用而言是难以满足的.

姿态和光照的变化对人脸图像的影响说到底是与人脸的 3D 结构有关的,假设人脸的 3D 形状、表面反射率已知或者可以估计,那么姿态和光照问题就可以更容易地解决.因此,一些基于模型的方法试图从人脸图像中将人脸内在属性(3D 形状和表面反射率)和外在成像条件(光照条件、绘制条件、摄像机参数等)分离开来,从而可以消除外在成像条件的影响而仅利用内在本质属性实现准确的分类识别.其中最著名的方法包括光照锥、对称 SFS 和 3D 形变模型方法.

Georghiades 提出了光照锥(illumination cone)方法^[6]来解决光照变化和姿态变化下的人脸识别.该方法能够根据给定的多幅(不少于 3 幅)相同姿态、不同光照条件的输入图像,估计出输入人脸的 3D 信息.其本质上是传统的光度立体视觉(photometric stereo)方法的一个变种,通过 SVD 迭代估计光照、人脸 3D 形状、表面点反射率,并最终利用人脸 3D 形状分布的先验知识(如对称性、鼻子为最近点等)作为约束求解人脸的 3D 形状.通常这种方法至少需要每个目标不同光照条件下的 7 幅图像,这对于大多数应用来说都太过苛刻,因此难以实用.与光照锥利用光度立体视觉方法不同,Zhao 则提出了采用从影调恢复形状(shape from shading,简称 SFS)方法进行人脸 3D 形状重建的思路,在传统 SFS 方法的基础上,利用了人脸的对称性先验知识,提出了 SSFS 方法(symmetric SFS)^[7].该方法只需要一幅输入人脸图像即可,但它需要通过其他方法估计输入图像的光照情况以及精确的对称轴信息,这都增加了该方法在实用上的困难.迄今为止,最成功的姿态和光照不变的人脸识别是 3D 变形模型方法(3D morphable model)^[8].该方法通过主成分分析对人脸的 3D 形状和纹理(表面反射率)分别进行统计建模,在此基础上建立了包含形状、纹理统计参数、Phone 模型参数、光照参数、摄像机内外参数、绘制参数等在内在的复杂成像模型,最终采用基于合成的分析(analysis-by-synthesis)技术通过优化算法估计这些参数,得到输入人脸的 3D 形状和纹理统计参数用于最终的分类识别.这种变形方法已经用于 FRVT2002^[11]中.遗憾的是,该方法需要求解一个涉及几百个参数的复杂连续优化问题,迭代优化过程耗费了大量的计算时间,整个拟合过程在一台 2GHzP4 处理器的工作站上,大约需要 4.5 分钟.这对于多数实用系统而言是难以忍受的.

基于上述分析,我们认为 3D 变形模型中采用的统计建模方法是利用人脸先验 3D 信息的最佳方式,所以本

文也采用了类似的建模方法,不同的是,为了避免 3D 形变模型中过分复杂的参数优化过程,我们没有采用稠密的统计模型,而仅仅采用了一个稀疏的关键特征点统计模型;同时也没有采用复杂的成像模型,而是采用了更为方便、实用的球面谐波商图像方法来处理光照估计和光照影响消除问题.这些措施极大地降低了算法的计算复杂度,使得整个处理过程可以在 1 秒内完成(P4 3.2G CPU 的机器上).当然,与 3D 形变模型相比,本文算法重建的人脸 3D 信息的精度会有较大的差距,但需要注意的是,我们的目标是实现对姿态、光照变化不敏感的人脸识别,并非精确的 3D 重建,而多数正面人脸识别系统都对较小的光照、姿态变化有一定的容忍能力,因此,本文在算法精度和速度方面进行折衷是合理的,本文的实验也表明了这一点.

本文第 1 节详细的描述我们提出的姿态和光照不变的人脸识别方法,第 1.1 节介绍基于 3D 稀疏变形模型的 3D 形状重构算法.第 1.2 节对基于球面谐波商图像的光照不变的纹理生成方法进行描述.本文算法的一些合成结果以及对姿态和光照不变的人脸识别实验结果在第 2 节给出.最后给出本文的结论.

1 姿态和光照不变的人脸识别

本文提出的光照、姿态不变的人脸识别系统的整体框架如图 1 所示.首先,对于给定的任意一幅人脸图像进行人脸检测和眼睛中心的定位,根据基于视图的子空间方法得到粗略的姿态类别,然后采用 ASM 或者 ESL 算法^[9]标定给定人脸图像的稀疏的关键特征点.基于 3D 稀疏变形模型,结合人脸图像的 2D 形状信息,重建对应于输入图像人脸的 3D 形状.借助于重建得到的特定人的 3D 形状,根据球面谐波商图像方法,将由输入人脸图像重新打光到预先定义的参考光照条件下,从而生成光照不变的纹理图像,即实现了参考光照条件下的 3D 人脸重建.进行了姿态和光照校正之后的人脸图像即可作为一般正面人脸识别系统的输入,与库存人脸图像进行匹配,得到识别结果.因此,我们的算法也可以看作是任何人脸识别系统的预处理步骤.

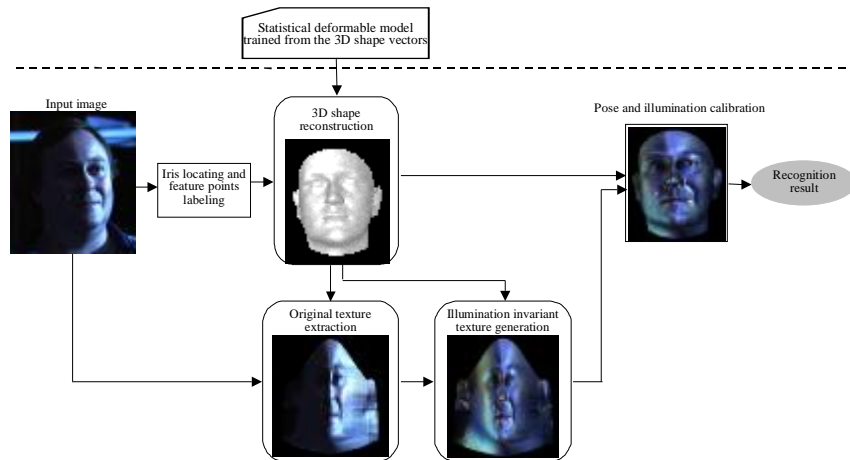


Fig.1 The framework of pose and illumination calibration for face recognition

图 1 用于人脸识别的姿态和光照校正的流程

下面,我们对上面所述框架中的两个关键问题分别进行描述,即基于 3D 稀疏变形模型的 3D 人脸形状重建和基于球面谐波商图像的光照不变的纹理生成.

1.1 基于单幅图像的 3D 人脸形状重构

由单幅人脸图像恢复出此特定人的 3D 结构是解决姿态问题的最直观、最本质的方法.然而,在没有任何假设的前提下,从单幅人脸图像恢复 3D 形状是一个病态的问题.文献[10]中明确指出,重建 3D 人脸所需图像的最小数目是 3.为克服需要多幅图像才能重建 3D 人脸形状这一困难,我们对 3D 人脸数据集中的形状向量进行主成分分析(principle component analysis,简称 PCA),建立统计变形模型.此模型即包含了人脸类的 3D 形状的先

验知识.

1.1.1 创建稀疏统计模型

本文采用 USF Human ID 3-D 数据库中的 100 个激光扫描的 3D 人脸作为训练集合创建统计变形模型^[11]. 所有这些人脸都被标准化到一个统一的标准的方向和位置. 人脸的形状由 75 972 个顶点表示, 为简化计算, 均匀下采样到 8 955 个顶点. 在本文算法中, 对形状和纹理按照不同的策略分开处理. 我们认为, 人脸的姿态只与一些关键特征点的位置有关, 而与图像的亮度无关. 通过将形状和纹理分开处理, 避免了复杂的优化过程, 节省了计算时间. 下面, 首先介绍稀疏统计模型的建立过程.

我们将人脸的 n 个顶点的 X, Y, Z 坐标串接起来, 组成形状向量来描述人脸的 3D 形状:

$$S = (X_1, Y_1, Z_1, \dots, X_n, Y_n, Z_n)^T \hat{\mathbf{I}} \hat{\mathbf{A}}^{3n}.$$

假设用于训练的 3D 人脸的数目是 m , 每个 3D 人脸向量可以描述为 S_i , 其中 $i=1, \dots, m$. 这些形状向量都已经 在尺度上进行了配准, 则任意一个新的 3D 人脸形状向量 S 都可表示为训练样本形状向量的线性组合形式, 即

$$S = \sum_{i=1}^m w_i S_i.$$

考虑到所有的人脸形状在整体上都是相似的, 只是不同的人脸形状之间存在一些小的差别. 由于 PCA 适合于捕获同类向量主成分方向上的变化, 并能够滤去各种测量噪声, 因此, 我们采用 PCA 对训练集合中的 3D 形状进行建模. 通过对协方差矩阵的特征值进行降序排列, 按序选择前 $d \ll (m-1)$ 个对应最大特征值的形状特征向量, 共同构成投影矩阵 P , 即形成统计变形模型:

$$S = \bar{S} + P\alpha \quad (1)$$

其中 \bar{S} 是平均的形状向量, α 是对应于 d 维投影矩阵 P 的系数向量. 当人脸发生姿态旋转时, 对上面的等式进行扩展. 对于 3D 形状中的任意一个顶点 $D=(x, y, z)^T$, 根据 PCA 特性, 由式(1)可知此式成立:

$$RD = R\bar{D} + RP_D \alpha.$$

其中, R 是由 3 个旋转角度表征的 3×3 维旋转矩阵, \bar{D} 是 \bar{S} 中对应此顶点的坐标值, P_D 是由 P 中抽取出的对应顶点 D 的 $3 \times d$ 维矩阵. 由此可知, 对于整个 3D 形状向量也同样满足上面的旋转变换. 我们引入符号 V^R 来表示这样的运算操作: 对一个 3D 形状向量 V 按旋转矩阵 R 进行旋转变换. 同样, P^R 表示对 P 中的每一列 3D 特征向量做同样的旋转变换. 因此, 将式(1)向多姿态情况扩展, 有下面的等式:

$$S^R = \bar{S}^R + P^R \alpha \quad (2)$$

有了这个扩展到多姿态的统计变形模型, 则任意姿态下的 3D 人脸形状都可以由变换到同姿态下的样本形状的 PCA 模型来表示. 但是仅仅给定一幅人脸图像, 如何利用人脸形状的先验知识来进行 3D 形状重建仍然是一个问题. 基于此, 我们进一步提出了 3D 稀疏变形模型(3D SDM). 目的是建立输入图像的关键特征点向量和 3D 稀疏变形模型之间的对应, 从而优化得到 3D 稀疏变形模型的系数向量. 我们认为此系数向量同样是密集统计变形模型的系数向量. 这样就可以很容易地实现对应于给定人脸的密集 3D 形状的重建.

与 3D 形状表示方法相似, 将 2D 人脸图像上的 k 个关键点的 X, Y 坐标串接起来, 表示为 S_I . 每一个 2D 的关键点在 3D 形状向量上都有且只有一个标号固定的点与其对应. 按照这种对应关系, 将 3D 形状上的对应于 2D 关键点的 3D 点抽取出来, 并同样连接起来表示, 标记为 S_C . 对平均形状和投影矩阵做同样的处理, 即形成稀疏的 \bar{S}_C 和 P_C . 由此得到 3D 变形模型的稀疏版本:

$$S_C = \bar{S}_C + P_C \alpha.$$

同样, 向多姿态情况进行扩展, 即创建得到 3D 稀疏变形模型(3D SDM):

$$(S_C)^R = (\bar{S}_C)^R + (P_C)^R \alpha.$$

1.1.2 基于 SDM 的 3D 重建

即使有了 3D 形状统计先验知识, 仅根据一幅 2D 图像恢复其 3D 形状仍然很困难. 本文的策略是以 PCA 由部分重构整体的特性为依据提出的. 我们试图根据输入图像的关键点信息以及相应的 3D 稀疏变形模型信息, 获取 PCA 模型的重构系数, 最后将优化的系数向量投影到完整的 PCA 模型上, 以得到对应输入图像人脸的密集

的 3D 形状向量.

我们接下来对稀疏的 3D 形状 $S\mathbf{c}$ 抽取其 X, Y 坐标, 形成 2D 形状分量, 标记为 $S\mathbf{c}$. 由于 $S\mathbf{c}$ 可以被认为是 3D 形状 S 的一部分, 因此, 下面的等式仍然近似成立:

$$S\mathbf{c} = \bar{S}\mathbf{c} + P\hat{\mathbf{a}}.$$

这里引入符号 V_f 来表示在 3D 形状 V 中按序抽取相应的 X, Y 坐标形成 2D 形状向量. ($\bar{S}\mathbf{c}_f$ 和 $(P\mathbf{c}_f$ 分别表示从稀疏的 3D 平均形状 $\bar{S}\mathbf{c}$ 和投影矩阵 $P\mathbf{c}$ 中提取的对应的 2D 关键点向量, 则符号 $(S\mathbf{c}_f$ 就表示从经过旋转矩阵 R 变换之后的稀疏 3D 形状向量 $S\mathbf{c}$ 中抽取 X, Y 坐标形成的 2D 形状向量. 因此, 任意姿态角度下的 3D 形状向量, 其对应 2D 关键点的稀疏向量的 X, Y 分量串联成的向量可以如下表示:

$$(S\mathbf{c}_f^a = \bar{S}\mathbf{c}_f^a + (P\mathbf{c}_f^a a \tag{3}$$

我们的目标是通过形状系数向量 a 来重构完整 3D 形状信息, 对应于特定人的形状系数向量可如下求解:

$$a = [(P\mathbf{c}_f^a)^+]^+ [S\mathbf{c}_f^a - (\bar{S}\mathbf{c}_f^a] \tag{4}$$

其中, $[(P\mathbf{c}_f^a)^+]$ 是伪逆转置矩阵, 可通过 $[(P\mathbf{c}_f^a)^+] = \left((P\mathbf{c}_f^a)^T (P\mathbf{c}_f^a) \right)^{-1} (P\mathbf{c}_f^a)^T$ 来计算. 此系数向量是根据 PCA 的部分向量求得的, 我们认为它同时也可作为完整 PCA 各分量对应的系数. 因此, 问题的关键就是根据图像关键特征点向量 S_i , 计算出向量 $(S\mathbf{c}_f^a \cdot S_i$ 和 $(S\mathbf{c}_f^a$ 之间的关系可以表示为

$$S_i = c[(S\mathbf{c}_f^a] + T \tag{5}$$

我们用对应于 x, y, z 三个坐标轴的旋转角度 a, b, g 来刻画旋转矩阵 R . 利用人脸图像上的 5 个关键特征点及其对应的 3D 模型 S 上的 5 个关键的 3D 点, 其组成的向量分别表示为 S_{1-5} 以及 $S\mathbf{c}$, 我们可以通过投影计算得到 3 个旋转角度参数. 用于计算旋转姿态参数的 5 个特征点是左、右眼中心, 鼻尖, 左、右嘴角.

下面, 我们将对此 3D 形状重建算法进行详细介绍. 迭代优化过程如下:

A. 通过对选择的 5 个关键特征点, 在输入人脸图像以及对应的 3D 形状上的点之间建立透视投影关系, $S_{i,s} = sQRSC_5 + (tx, ty)^T$ 联立方程组, 求解计算得到旋转矩阵 R 的参数. 其中, Q 为 2×3 的投影矩阵, 即 $Q = \begin{bmatrix} \frac{af}{s} & 0 & 0 \\ \frac{bf}{s} & 1 & 0 \end{bmatrix}$, 仅获取 x, y 方向的信息用来计算. s 是尺度因子, tx, ty 分别是 x, y 方向平移分量.

B. 在 S_i 和 $(S\mathbf{c}_f^a$ 两个向量对应的位移向量 T 和尺度变化因子 c 通过下面等式计算: 尺度因子为

$$c = \frac{\dot{\mathbf{a}} \left((X_i^t - \bar{N}X_0)X_i + (Y_i^t - \bar{N}Y_0)Y_i \right)}{\dot{\mathbf{a}} \left((X_i^t)^2 + (Y_i^t)^2 \right)}.$$

然后, 更新位移向量分量:

$$\bar{N}X = \frac{1}{k} \dot{\mathbf{a}} \left(X_i^t - c_0 X_i \right), \quad \bar{N}Y = \frac{1}{k} \dot{\mathbf{a}} \left(Y_i^t - c_0 Y_i \right),$$

位移向量为

$$T = (\bar{N}X, \bar{N}Y, \bar{N}X, \dots, \bar{N}Y)^T.$$

其中, $\bar{N}X_0$ 和 $\bar{N}Y_0$ 分别是 $\bar{N}X$ 和 $\bar{N}Y$ 在前一次迭代计算得到的值, 在第 1 次计算时被设置为 A 步骤中求解得到的 tx 和 ty . X_i^t 和 Y_i^t 是输入人脸图像中的特征点的坐标值. X_i 和 Y_i 是 $(S\mathbf{c}_f^a$ 中对应的点的 x, y 方向分量值. 重复 B 步骤 2~3 次即可计算出恰当的位移和尺度因子.

C. 利用 B 中得到的 T 和 c , 通过等式(5)更新 $(S\mathbf{c}_f^a$.

D. 得到了新的 $(S\mathbf{c}_f^a$, 我们就可以很容易地根据等式(4)来计算形状系数向量 a .

E. 根据等式 $S\mathbf{c} = \bar{S}\mathbf{c} + P\hat{\mathbf{a}}$, 我们可以重构出对应于特定人的稀疏的 3D 人脸形状 $S\mathbf{c}$.

F. 重复步骤 A~E, 直到形状系数向量收敛.

最终,由密集统计变形模型(等式(1)),我们可以重构得到对应于输入人脸的 3D 形状.为得到更加精细的 3D 形状的解,我们按照给定 2D 图像上的特征点的坐标来对 3D 形状顶点进行进一步调整.

一旦得到了人脸精细的 3D 形状信息,再结合此特定人的纹理图像,对姿态的校正就可以简单地通过旋转 3D 人脸模型来实现了.

1.2 基于球面谐波图像的光照不变的纹理生成

根据上节中恢复得到的 3D 形状和姿态参数,人脸区域可以直接从给定的 2D 人脸图像中抽取得到.然而,所得到的人脸图像并不是真正的纹理图像,而是随着光照的变化而变化的.虽然生成特定人本质的纹理是很直观的想法,但是本文并不直接恢复纹理信息,而是转而采取去光的策略来消除光照的影响,将提取的人脸区域的光照条件校正到参考光照条件下,即形成光照不变的纹理图像^[12].最终,这个标准参考光照下的纹理可以被绘制到 3D 人脸形状上,从而重建出完整的与光照无关的 3D 人脸.

由于反射等式可以被看作是一个卷积,因此很自然地在频域空间来对其进行分析.对于球面谐波,Basri 等人^[13]证明了亮度的绝大部分能量都限制在 3 个低阶的部分,其频域公式为

$$E(\mathbf{a}, \mathbf{b}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l E_{lm} Y_{lm}(\mathbf{a}, \mathbf{b}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l A_l L_{lm} Y_{lm}(\mathbf{a}, \mathbf{b}) \gg \sum_{l=0}^2 \sum_{m=-l}^l A_l L_{lm} Y_{lm}(\mathbf{a}, \mathbf{b}) \quad (6)$$

其中, $A_l (A_0 = \pi, A_1 = 2\pi/3, A_2 = \pi/4)$ ^[13]是 Lambertian 反射的球面谐波系数, L_{lm} 是入射光线的系数, Y_{lm} 是球面谐波函数.

给定一幅纹理图像 I ,对于每一个像素 (x,y) ,下面等式几乎都成立:

$$I(x, y) = \mathbf{r}(x, y)E(\mathbf{a}(x, y), \mathbf{b}(x, y)).$$

这里, $\mathbf{a}(x, y)$ 和 $\mathbf{b}(x, y)$ 可由 3D 人脸形状的法线向量计算得到.假设人脸的反照率 \mathbf{r} 为常数, $E_{lm} = A_l Y_{lm}$ 表示谐波图像, E 是 E_{lm} 的 $n \times 9$ 维矩阵.其中, n 是纹理图像的像素总数,则光照条件 L 的系数就可以通过最小二乘解得:

$$\hat{L} = \arg \min_L \|E(\mathbf{r}L) - I\| \quad (7)$$

如果我们已经估计得到给定图像的光照条件,就可以很容易地将其重新打光到标准光照下.对图像上 (x,y) 处一个确定的点 P ,它的法线为 (\mathbf{a}, \mathbf{b}) ,反照率为 $\mathbf{r}(x, y)$,则原始图像上和光照校正后 P 点的亮度分别是:

$$I_{org}(x, y) = \mathbf{r}(x, y) \sum_{l=0}^2 \sum_{m=-l}^l A_l \hat{L}_{lm} Y_{lm}(\mathbf{a}, \mathbf{b}) \quad (8)$$

$$I_{can}(x, y) = \mathbf{r}(x, y) \sum_{l=0}^2 \sum_{m=-l}^l A_l L_{lm}^{can} Y_{lm}(\mathbf{a}, \mathbf{b})$$

两种不同光照下的商图像定义为

$$R(x, y) = \frac{I_{can}(x, y)}{I_{org}(x, y)} = \frac{\mathbf{r}(x, y) \sum_{l=0}^2 \sum_{m=-l}^l A_l L_{lm}^{can} Y_{lm}(\mathbf{a}, \mathbf{b})}{\mathbf{r}(x, y) \sum_{l=0}^2 \sum_{m=-l}^l A_l \hat{L}_{lm} Y_{lm}(\mathbf{a}, \mathbf{b})} = \frac{\sum_{l=0}^2 \sum_{m=-l}^l A_l L_{lm}^{can} Y_{lm}(\mathbf{a}, \mathbf{b})}{\sum_{l=0}^2 \sum_{m=-l}^l A_l \hat{L}_{lm} Y_{lm}(\mathbf{a}, \mathbf{b})} \quad (9)$$

由于给定图像的光照条件和参考光照条件都是确定的,因此对于给定人来说,参考光照条件相对于原始光照条件下的商图像就已经确定了.

进而,基于原图像和商图像,光照校正后的纹理图像可以如下计算得到:

$$I_{can}(x, y) = R(x, y) \cdot I_{org}(x, y) \quad (10)$$

得到了精细的 3D 形状和去除光照影响的纹理后,我们就实现了根据输入的单幅任意光照下的非正面图像重构出对应于特定人的标准光照条件下的 3D 人脸.对于纹理图像上不可见的点,采取插值策略进行填充.

2 实验与分析

在这一节,我们通过姿态和光照不变的人脸识别来评价本文提出的姿态和光照校正算法的性能.对于一幅任意光照下的非正面人脸图像,重建它的与光照无关的 3D 人脸,即无论输入图像的光照条件如何,最终重构的

3D 人脸对应的纹理都被校正到预先定义的参考光照条件下.姿态的校正是通过将 3D 人脸旋转到预先定义的标准姿态下实现的.经过校正后的人脸图像被作为一般人脸识别系统的输入来进行识别.

2.1 姿态不变的人脸识别实验结果

首先,对仅仅是姿态发生变化的情况进行测试,在 CMU PIE 数据库^[14]中对 4 个姿态子集进行测试,分别是姿态集合 05(右转 22.5°)、姿态集合 29(左转 22.5°)、姿态集合 37(右转 45°)和姿态集合 11(左转 45°).Gallery 中的图像来自姿态集合 2,其中都是正面图像.本节仅仅对姿态变化情况进行测试,假设给定的图像其光照条件是均衡的.因此,不必对光照进行特别处理,即用从图像中提取出来的人脸区域来近似标准光照条件下的纹理图像,然后进行 3D 人脸重建,实现姿态校正.

我们采用的人脸识别方法是 Gabor PCA 加 LDA 方法,其思想与 GFC 相似^[15].训练图像从 CAS-PEAL 数据库^[16]中选取,共 300 个人,平均每人选择 6 个姿态图像,10 个正面图像.由于 CAS-PEAL 库采集条件与 CMU PIE 库相距甚远,因此,避免了训练图像和测试图像之间条件相似给识别结果带来的影响.在此实验过程中,特征点是手工标定的.

对于姿态校正策略进行人脸识别,其最重要的一点就是校正后的视图一定要在外观上与 Gallery 中的原型视图很相似.我们在图 2 里给出了一些基于 3D 人脸重建的姿态校正后的结果,用来进行视觉上的评价.图 2 中第 1 行是原始的 4 个非正面姿态下加了掩模的图像,第 2 行是对应的姿态校正后的图像,校正视图右侧的一列是来自姿态集合 27 的 gallery 图像,用作校正视图的参考.可见,对 4 个姿态的人脸进行姿态校正,校正后的人脸与 Gallery 中原始正面人脸图像极为相似.并且,在 P4 处理器,主频 3.2GHz 的电脑上,本文的方法进行一次完整的人脸重建只需 1 秒的时间.与文献[8]中的 3D 变形模型方法相比,计算复杂度大大降低.3D 变形模型方法的一次拟合过程在一台 2GHz,P4 处理器的工作站上,大约需要 4.5 分钟^[8].我们对这 4 个姿态子集的图像进行姿态校正,继而进行人脸识别,识别结果在表 1 中列出.校正之后,在 4 个姿态集合上的累积识别率比用原始人脸图像进行识别的情况有很大程度上的提高,首选识别率平均达到 94.85%.我们还将此结果与特征光场方法(eigen light field)的结果进行了比较.特征光场方法可以用来解决人脸识别中的姿态和光照问题^[17].该方法通常采用两种不同的标准化策略,分别是 3 点标准化和多点标准化.我们的结果与这两种标准化策略的结果进行比较,识别率都有提高(见表 1).实验结果表明,基于 3D 人脸重建的姿态校正方法具有很好的性能.因此,我们认为在计算时间降低的情况下,基于本文方法进行的姿态不变的人脸识别仍然取得了另人满意的识别率.

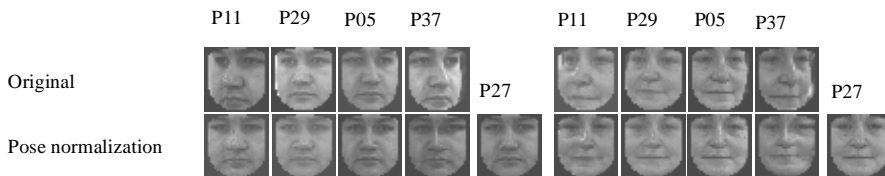


Fig.2 The pose normalized images

图 2 姿态校正的结果

由于这里用于 3D 人脸重建的特征点是手工标定的,而各种多姿态特征点定位方法在定位人脸关键点的过程中都不可避免地存在误差,因此为了衡量本文算法对特征点定位误差的鲁棒性,我们人为地对手工标定的特征点加高斯噪声.基于有高斯误差的特征点定位结果,我们重新进行 3D 人脸重建、姿态校正以及后续的人脸识别,从而验证定位的误差对使用本文策略进行姿态不变的人脸识别结果的影响.这里为尽量减少其他因素的影响,精确地分析特征点定位误差对识别结果的影响,我们采用简单的计算相关性的方法作为识别的度量.我们还尝试了对手工标定的人脸关键特征点增加 5 组不同的噪声,其均值为 0,方差分别为 1.0,1.5,2.0,2.5 以及 3.0.对特征点定位结果加入高斯噪声的姿态不变人脸识别结果见表 2.我们发现,对特征点定位结果增加不同程度的高斯扰动,并没有造成识别结果发生较大的变化.

Table 1 The recognition results on the original images and the pose normalized images in the 4 different pose sets of CMU-PIE database with Gabor PCA add LDA recognition strategy and the comparison with Eigen Light Fields

表 1 在 CMU PIE 数据库中 4 个姿态子集上使用 Gabor PCA + LDA 识别策略分别对原始图像和姿态校正图像进行识别的结果,以及与特征光场方法的结果比较

| Method | PCA+LDA with original images (%) | GPCA+LDA with normalized images (%) | ELF method ^[17] (3-Points normalization) (%) | ELF method ^[17] (Multi-Points normalization) (%) |
|-------------|----------------------------------|-------------------------------------|---|---|
| Pose set 29 | 82.3 | 97.1 | 86 | 56 |
| Pose set 05 | 97.1 | 98.5 | 88 | 94 |
| Pose set 11 | 35.3 | 95.6 | 76 | 88 |
| Pose set 37 | 55.9 | 88.2 | 74 | 88 |

Table 2 Recognition performance comparison between using the original landmarks labeled manually and the imprecise landmarks added Gaussian noises of CMU PIE database with correlation recognition strategy

表 2 在 CMU PIE 数据库姿态子集上,采用相关性识别策略,对基于手工标定结果与增加高斯噪声的特征点结果进行姿态不变的人脸识别的性能比较

| Recognition rate | Landmarks labeling manually (%) | Landmarks added Gaussian noises (%) | | | | |
|------------------|---------------------------------|-------------------------------------|-------------|-------------|-------------|-------------|
| | | (u=0;g=1.0) | (u=0;g=1.5) | (u=0;g=2.0) | (u=0;g=2.5) | (u=0;g=3.0) |
| Pose set 29 | 63.24 | 64.71 | 64.71 | 64.71 | 63.24 | 60.29 |
| Pose set 05 | 76.47 | 76.47 | 75.00 | 75.00 | 76.47 | 77.94 |
| Pose set 11 | 54.41 | 54.41 | 51.47 | 51.47 | 57.35 | 55.88 |
| Pose set 37 | 58.82 | 58.82 | 60.29 | 60.29 | 58.82 | 60.29 |

2.2 姿态和光照不变的人脸识别实验结果

我们用本文提出的策略对姿态和光照同时发生变化的人脸图像进行校正,对其校正结果进行识别实验.在此实验中,使用 CMU PIE 的“illum”集合中的部分图像.该集合提供了姿态和光照进行严格控制的多个姿态下人脸的多种光照条件下的视图.我们对 05 和 29 这两个姿态子集共 2 856 幅图像进行测试,每个姿态子集包含 21 种不同的光照条件,其闪光灯序列号为 02-21.将闪光灯号“11”下的姿态 27 的正面图像作为 gallery,其他的测试图像就都被校正到正面姿态下,标准光照“11”下.

在图 3 中,我们给出了一些姿态和光照标准化后的图像示例(以姿态集合 05 中的一个样本为例),图像(a)中是按照 02-21 光照顺序进行排列的.我们可以发现,经过校正后的图像(b)与 galley 的样本图像(c)更为相似,而图像(a)中的原始图像在姿态与光照条件上都与图像(c)存在较大的差别,给识别带来了很大的困难.

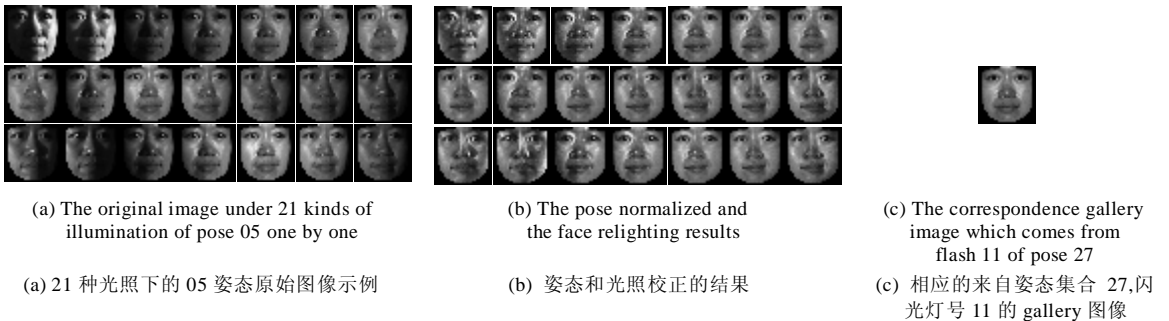


Fig.3 The pose and illumination calibrated results
图 3 姿态和光照校正结果

本文提出的姿态和光照的校正可作为任意人脸识别系统的预处理步骤,因此在本实验中我们只使用简单的相似度计算,根据最近邻匹配策略进行识别实验.人脸识别的实验结果在图 4 中给出.经过本文的姿态和光照校正,使人脸识别系统的识别率有了显著提高,平均约提高了 48.4%.对此结果进行分析,本策略对几个集合的识别率提高得不是很理想,如 02,03,12,13 这 4 个光照集合.由图 3 我们可以看到,这 4 个集合包含的阴影区域都很大.而球面谐波函数对解决这种阴影问题有其自身的局限性,因此使得经过校正处理之后的识别率虽然与原来相比有了很大程度的提高,但是仍然不能令人满意.如果去除这 4 个阴影过于严重的集合,我们仅仅使用相似度

匹配的策略来进行识别,经过校正之后的识别率也可以达到 92.5%左右,比未经校正情况提高了约 56.75%。因此,即使存在着一些限制,本文的方法用来解决人脸识别中的姿态、光照问题仍然是很有效的。

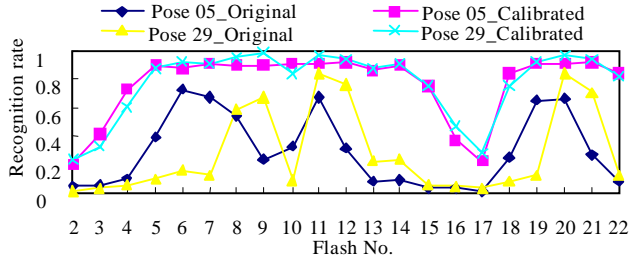


Fig.4 Recognition results on 2 pose subsets under 21 different lightings in CMU PIE Database with the correlation match strategy

图 4 根据相关系数匹配策略进行姿态和光照校正前后,对 CMU PIE 21 种光照条件下的两个姿态子集识别结果

3 结论与后续工作

针对人脸识别中的姿态和光照问题,本文提出了一种基于 3D 人脸重建进行姿态校正并将原图像重新加光到标准光照条件下的方法。作为一种基于模型的方法,我们采用 3D 统计模型作为先验约束,通过形变模型匹配快速重建 3D 人脸模型,然后利用重建的 3D 人脸模型就可以很容易地实现姿态的校正,并利用球面谐波光照处理方法对其重新打光,从而可以实现姿态和光照的归一化,极大地提高了现有人脸识别系统对姿态、光照的适应能力。在 PIE 光照、姿态人脸库上的实验表明,该方法能够有效地改善现有识别系统的识别性能。

虽然与更为完备的 3D 形变模型相比,本文的 3D 重建方法的精度是有限的,但是,为了进行姿态和光照的归一化,我们并不需要完全精确的 3D 人脸模型就可以取得令人满意的结果,更关键的是,本文所提出的方法建模速度得以提高,能够在可以接受的短时间内完成整个处理过程,从而可以满足很多应用系统的实际需求,如公安系统可以使用的照片比对系统。

由于球面谐波函数不能很好地对阴影进行建模,因此,我们的工作重点将是研究如何克服阴影区域的影响,进行阴影的消除,以期能够更好地解决姿态和光照问题。

References:

- [1] Phillips PJ, Grother P, Micheals RJ, Blackburn DM, Tabassi E, Bone M. Face recognition vendor test 2002: Evaluation report. Technical Report, NISTIR 6965, Gaithersburg. National Institute of Standards and Technology, 2003.
- [2] Turk M, Pentland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1991,3(1):71- 86.
- [3] Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1997,19(7):711- 720.
- [4] Gross R, Matthews I, Baker S. Eigen light-fields and face recognition across pose. In: Williams AD, ed. *Proc. of the Conf. on FGR 2002*. Washington: IEEE Computer Society, 2002. 3- 9.
- [5] Zhou S, Chellappa R. Illuminating light field: Image-Based face recognition across illuminations and poses. In: Azada D, ed. *Proc. of the FGR 2004*. Washington: IEEE Computer Society, 2004. 229- 234.
- [6] Georgiades AS, Belhumeur PN, Keigman DJ. From few to many: Illumination cone models for face recognition under variable lighting and poses. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2001,23(6):643- 660.
- [7] Zhao W, Chellappa R. SFS based view synthesis for robust face recognition. In: Crowley JL, ed. *Proc. of the FGR 2000*. Washington: IEEE Computer Society, 2000. 285- 292.
- [8] Blanz V, Vetter T. Face recognition based on fitting a 3D morphable Model. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003,25(9):1063- 1074.

- [9] Chai X, Shan S, Gao W, Chen X. Example-Based learning for automatic face alignment. Journal of Software, 2005,16(5):718- 726 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/718.htm>
- [10] Huang T, Lee C. Motion and structure from orthographic projections. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1989,11(5):536- 540.
- [11] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces. In: Rockwood A, ed. Proc. of the SIGGRAPH'99. New York: ACM Press, 1999. 187- 194.
- [12] Qing L, Shan S, Gao W, Du B. Face recognition under generic illumination based on harmonic relighting. Int'l Journal of Pattern Recognition and Artificial Intelligence, 2005,19(4):513- 531.
- [13] Basria R, Jacobs D. Lambertian reflectance and linear subspaces. In: Bob W, ed. Proc. of the ICCV 2001. Washington: IEEE Computer Society, 2001. 383- 390.
- [14] Sim T, Baker S, Bsat M. The CMU pose, illumination, and expression (PIE) database. In: Williams AD, ed. Proc. of the Conf. on FGR 2002. Washington: IEEE Computer Society, 2002. 46- 51.
- [15] Liu C, Wechsler H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans. on Image Processing, 2002,11(4):467- 476.
- [16] Gao W, Cao B, Shan S, Zhou D, Zhang X, Zhao D. The CAS-PEAL large-scale Chinese face dataBase and evaluation protocols. Technique Report, No.JDL-TR_04_FR_001, Beijing: Joint Research & Development Laboratory, the Chinese Academy of Sciences, 2004.
- [17] Gross R, Matthews I, Baker S. Appearance-Based face recognition and light-fields. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2004,26(4):449- 465.

附中文参考文献:

- [9] 柴秀娟,山世光,高文,陈熙霖.基于样例学习的面部特征自动标定算法.软件学报,2005,16(5):718- 726. <http://www.jos.org.cn/1000-9825/16/718.htm>



柴秀娟(1978—),女,黑龙江木兰人,博士生,主要研究领域为图像图形处理,模式识别,多姿态人脸识别.



陈熙霖(1965—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为计算机视觉,模式识别,人工智能..



山世光(1975—),男,博士,助理研究员,主要研究领域为模式分析,机器智能.



高文(1956—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为模式识别,人工智能,计算机视觉,视音频编码.



卿来云(1974—),女,博士,主要研究领域为图像处理,模式识别.